Least Squares Consistent Estimates for Arbitrary Regression Functions over an Abstract Space

Silvano Fiorin
Department of Statistical Sciences
University of Padua

Francesca Parpinel
Department of Economics
Ca' Foscari University of Venice

January 30, 2011

Corresponding author: Silvano Fiorin, Department of Statistical Sciences University of Padua, v. C.Battisti, 241, 35121 Padova, Italy
email: `fiorin@stat.unipd.it`

Short title: LS Consistent Estimates for Regression Functions

Abstract

In this paper we propose a new approach to sieve estimation for a general regression function when the dimension of the finite dimensional subspaces is a random quantity depending on the values of the observations.

The technique is introduced with the help of a simulation study on a functional linear model under extremely mild assumptions.

A sketch of the proof concerning the main statements is then given in the more general case when the regression function is not necessarily linear.

1 Preliminaries

In this work we study the problem of estimating a general regression function in the two different contexts which will be referred to as the *functional linear model* and the *general regression model*.

The functional linear model (FLM) combines a scalar random variable $Y \in \mathbb{R}$ and a random function X with finite second moment taking values in $L^2[0,T]$, i.e. $E\|X\|^2 = \int_0^T E|X(t)|^2 dt < \infty$, by the equality

$$Y = \int_0^T \theta_0(t)X(t)dt + \epsilon$$

where the assigned function $\theta_0 \in L^2[0,T]$ is called the regression function and the error term ϵ is zero mean and uncorrelated with X, i.e. $E(X(t)\epsilon)^2 = 0$, $\forall t \in [0,T]$.

The FLM is a well known topic in Statistics and many applications have been developed in different areas, such as chemistry by Frank and Friedman (1993), finance by Preda and Saporta (2005) and climatology by Besse et al. (2000). From a statistical point of view, several techniques have been developed to estimate the unknown regression function θ_0.

Partial least squares and principal components regression were adopted for estimation by Frank and Friedman (1993); an estimator was obtained by Cardot et al. (1999) using spectral analysis of the empirical second moment operator of X, and splines estimators were obtained by Cardot et al. (2003) and Cardot et al. (2007). Cardot and

Johannes (2010) considered a threshold rule for the estimation and Reiss and Ogden (2007) interestingly compare several methods, including functional component regression. Furthermore estimation for generalized functional linear models was proposed by Müller and Stadtmüller (2005).

The general regression model (GRM) deals with a couple (X, Y) where X denotes a random element taking values into an arbitrary complete and separable metric space $\mathcal{X}$, having the Borel σ-field $\mathcal{B}_{\mathcal{X}}$, and Y denotes a real random variable such that

$$E(Y|X = x) = T_0(x), \quad \text{where } T_0 \in L^2(\mathcal{X}, \mathcal{B}_{\mathcal{X}} P_X)$$

is an assigned square integrable function belonging to the separable Hilbert space L^2 and P_X denotes the probability measure induced by X.

As stated above, in the FLM, several techniques were produced to estimate the regression function θ_0 belonging to an infinite dimensional vector space, through a sequence of finite dimensional vector subspaces $\mathcal{S}_{m(n)}$ whose dimension $m(n)$ depends on the sample size n. In all cases an estimate $\hat{\theta}_n \in \mathcal{S}_{m(n)}$ is derived adopting some strategy and the consistency is reached if the dimension $m(n)$ tends to infinity *slowly enough* when n is divergent to infinity.

Our analysis falls within the above mentioned framework, although we consider a more simplified assumptions system. For example, it may be easily checked that the choice of the subspace dimension, $m(n)$, is typically chosen on the basis of the sample size n (for instance Cardot et al., 1999, put $m(n) = o(\log(n))$) with no regard to other relevant components. Our point of view may be described by the following question: Is it reasonable to use the same dimension $m(n)$ of the subspace if θ_0 is either a very smooth periodic function or a discontinuous one with unbounded derivatives?

A second problem arises considering the assumptions under which the strong consistency of estimators is proved; the introduced conditions generally define restrictive hypotheses and the proposed simulations concern estimates for regular and smooth functions. Thus it is reasonble to ask: What happens if we check an estimation method using a function θ_0 which is not regular?

The above questions led us to adopt a least squares strategy we introduce here in the case of a FLM, but we may easily generalize in the case of a GRM; in fact we provide the proof of some statements in the more general case of a GRM.

Our estimation procedure is as follows: denoting by $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$ the set of n observations and introducing the orthonormal base $\{b_j : j \geq 1\}$ for the Hilbert space $L^2[0, T]$, for each fixed $m = 1, 2, \ldots, n$ let us consider the finite dimensional space $\mathcal{S}_m(b_j : j = 1, \ldots, m)$ generated by the first m elements of the orthonormal base, thus we denote by $\hat{\theta}_n^m$ the global minimizer for the random function

$$L_n(a_j, j = 1, \ldots, m) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{m} a_j \int_0^T b_j(t) x_i(t) dt \right)^2$$

over the subspace $\mathcal{S}_m$. The estimation procedure consists of a rule that selects an element, denoted by $\hat{\theta}_n^{\bar{m}_n}$, within the class of functions $\Theta_n = \{\hat{\theta}_n^m : m = 1, 2, \ldots, n\}$ on the basis of the distances $\{||\hat{\theta}_n^m - \hat{\theta}_n^{m-1}|| \ \forall m = 2, \ldots, n\}$. The estimation is now obtained taking $\hat{\theta}_n = \hat{\theta}_n^{\bar{m}_n}$ where $\bar{m}_n$ is the dimension of the subspace selected by the procedure.

Several properties of such a method may be investigated through simulations. Two different kinds of regression functions are considered:

i. the regular smooth and periodic function $\theta_0(t) = \sin(4\pi t)$, $\forall t \in [0, \pi]$ (as in Cardot et al., 1999);

ii. the discontinuous unbounded function with unbounded derivatives $\theta_0(t) = \log|t-1|$, $\forall t \in [0, \pi]$.

The strong difference between i. and ii., in terms of regularity properties, will produce a meaningful effect on the dimension $\bar{m}_n$ of the selected subspace. The simulations were performed on the basis of $n = 200$ observations where the regressor X is a Brownian motion; if $\hat{\theta}^{m^*}_{200}$ denotes the set of the global minimizers for the class of distances $\{||\hat{\theta}^{m}_{200} - \theta_0|| : \forall m = 1, 2, \ldots, 200\}$, in both the cases i. and ii. we observe that the proposed estimates are very close to the global minimizers $\hat{\theta}^{m^*}_{200}$.

In fact, in case i. we get $\hat{\theta}^{m^*}_{200} = \{\hat{\theta}^{m}_{200} : m = 23, 24, 25, 26, 27, 28, 29, 30, 31\}$ (see Table 3) where $\{||\hat{\theta}^{m}_{200} - \theta_0|| = 0.04 \ \forall m = 23, \ldots, 31\}$ and $\hat{\theta}^{\bar{m}_{200}}_{200}$ is any element belonging to the set $\{\hat{\theta}^{m}_{200} : m = 20, 21, 23, 24, 25, 26, 28, 29, 30\}$ where the maximum of the distances from θ_0 is reached if $m = 20, 21$ giving $||\hat{\theta}^{20}_{200} - \theta_0|| = ||\hat{\theta}^{21}_{200} - \theta_0|| = 0.05$.

As far as case ii. is concerned, we have $\hat{\theta}^{m^*}_{200} = \hat{\theta}^{8}_{200}$ with $||\hat{\theta}^{8}_{200} - \theta_0|| = 0.39$ and the estimation procedure gives $\hat{\theta}_{200} = \hat{\theta}^{\bar{m}_{200}}_{200} = \hat{\theta}^{5}_{200}$ with $||\hat{\theta}^{5}_{200} - \theta_0|| = 0.56$.

Thus the two choices for θ_0 produce a big difference on the dimension $\bar{m}_{200}$ of the subspace for the estimates. In fact in the regular case $\bar{m}_{200}$ is any value included in the set $\{20, 21, 23, 24, 25, 26, 28, 29, 30\}$, while in the irregular case $\bar{m}_{200} = 5$. In both cases the estimates are very close to the optimal choice $\hat{\theta}^{m^*}_{200}$ meaning that the dimension of the subspace is strongly dependent on regularity properties of the regression function to be estimated. Furthermore, case ii. shows that the good estimates have a subspace dimension satisfying the condition $5 \leq m_n \leq 10$ whereas outside this context the distance from θ_0 increases very quickly (see the third column of Table 1).

This implies that the subspace dimension needs a good level of accuracy when regu-

larity properties of θ_0 are relaxed; thus, if θ_0 is really unknown, the choice for m_n based only on the sample size n is a too raw criterion and the only information concerning θ_0 is available in the observation (x_i, y_i). This is the reason why we assume that the subspace dimension depends on the observations.

The assumptions introduced in our approach are a relevant argument too; Section 5 deals with some technical results and proofs dealing with the GRM, where only the assumptions A1 and A2 are adopted in order to ensure the strong consistency of estimates. In the case of a FLM these not restrictive conditions may be rewritten in a simplified version where the main hypothesis is introduced using the error random variable: indeed we require that ϵ is a real random variable with finite variance and bounded density function. No assumptions are needed for the regressor X.

Finally some specifications concerning this paper may be useful. All the estimation procedure is built through an example where the regressor is a Brownian motion, following the framework provided by the relevant scholarship supplied above. In any case, using this example does not affect the generality of the method: indeed considering Brownian motion as a regressor X is not strictly required by our assumptions set.

2 Introduction

In order to introduce the problems studied in this paper and the adopted approach, we first provide a simulation result from an estimation procedure following regression model

$$Y = \int_0^\pi \theta_0(t)X(t)dt + \epsilon, \tag{1}$$

where $\{X(t), t \in [0, \pi]\}$ is a Brownian motion, ϵ is a standard Gaussian random variable and $\theta_0 \in L^2([0, \pi], dt)$ is an assigned function to be estimated. The simulation results are

obtained using a sieve-type least squares estimation technique: for example, if we let θ_0 be the global unique minimizer of the strictly convex function

$$L(\theta) = E\left[\left(Y - \int_0^\pi \theta(t)X(t)dt\right)^2\right] \qquad \forall \theta \in L^2([0,\pi], dt),$$

an estimation can be derived for θ_0 on the basis of the n observations $\{(x_i(\cdot), y_i) : i = 1, 2, \ldots, n\}$, where $x_i(\cdot)$ denotes the i-th observed trajectory for the regressor X. Let $\hat{\theta}_n$ denote the global minimizer of the empirical function

$$L_n(a_j, j = 1, \ldots, m_n) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{m_n} a_j \int_0^\pi \frac{\sin(jt)}{\sqrt{\frac{\pi}{2}}} x_i(t)dt\right)^2, \tag{2}$$

which is defined over the m_n-dimensional subspace $\mathcal{S}_{m_n} = Sp\left\{\frac{\sin(jt)}{\sqrt{\pi/2}} : j = i, \ldots, m_n\right\}$ generated by the first m_n elements of the orthonormal base $\left\{\frac{\sin(jt)}{\sqrt{\frac{\pi}{2}}} : j \geq 1\right\}$. Then the strong consistency of the $\hat{\theta}_n$ obtained via the sieves method holds if the dimension m_n of the subspace $\mathcal{S}_{m_n}$ increases *slowly enough* to infinity when the number n of observations diverges. Usually m_n is a deterministic quantity which depends on n and tends to infinity at a given rate.

The approach we adopt is based on a different *policy* about the subspaces dimension m_n which is a random quantity depending on the observations $\{(x_i(\cdot), y_i) : i = 1, 2, \ldots, n\}$.

In Section 3 we give a description of our estimation procedure by means of the results of a simulation study.

3 Definitions and background

This section introduces the definitions and the concepts needed to construct the estimates $\hat{\theta}_n$. In order to explain the intuitive meaning and the reasoning underlying our procedure,

from an operative point of view, we anticipate some theoretical results, here denoted as *Statements*, formally proved in Section 5.

The positive quantity

$$L(\theta) = E\left[\left(Y - \int_0^{\pi} \theta(t)X(t)dt\right)^2\right] \qquad \forall \theta \in L^2([0,\pi], dt) \tag{3}$$

defines a strictly convex function having the unknown θ_0 as its unique global minimizer. Denoting by n the number of available observations $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$, as we prefer the simplified notation x_i rather than $x_i(\cdot)$ to denote the observed trajectories of the regressor, we take the orthonormal base

$$\left\{\frac{\sin(jt)}{\sqrt{\frac{\pi}{2}}} : j \geq 1\right\} \tag{4}$$

for the Hilbert space $L^2([0,\pi], dt)$ and then we compute the global minimizers $\hat{\theta}_n^m$ (for each $m = 1, 2, \ldots, n$) of the empirical function

$$L_n(a_j, j = 1, \ldots, m) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{m} a_j \int_0^{\pi} \frac{\sin(jt)}{\sqrt{\frac{\pi}{2}}} x_i(t)dt\right)^2, \tag{5}$$

which is defined over the m-dimensional subspace

$$\mathcal{S}_m = Sp\left\{\frac{\sin(jt)}{\sqrt{\pi/2}} : j = i, \ldots, m\right\} = \left\{\sum_{j=1}^{m} a_j \frac{\sin(jt)}{\sqrt{\frac{\pi}{2}}} : (a_j : j = 1, \ldots, m) \in R^m\right\} \tag{6}$$

generated by the first m elements of the orthonormal base. Thus we obtain the set of Least Square (LS in short) finite dimensional estimates

$$\left\{\hat{\theta}_n^m : m = 1, \ldots, n\right\} \tag{7}$$

where m denotes the dimension of the subspace $\mathcal{S}_m$ and n the number of observations. The estimation technique described below consists of a rule allowing to choose an element denoted by $\hat{\theta}_n$ within the set (7). The simulation study is based on $n = 200$

observations, when $\theta_0(t) = \ln|t-1|, \forall t \in [0,\pi]$, used to compute the LS estimates $\left\{\hat{\theta}^m_{200} : m = 1,\ldots,200\right\}$. As an example we give below in Table 1 the results of one replication of the experiment. The table is divided into three blocks each of three columns. The first one contains the subspace dimension $m = 1,\ldots,200$; the second one provides the distances

$$\left\{||\hat{\theta}^m_{200} - \hat{\theta}^{m-1}_{200}|| : m = 1,2,\ldots,200\right\} \tag{8}$$

in term of L^2 norm $||\theta|| = \left(\int_0^\pi \theta^2(t)dt\right)^{1/2}$. When $m = 1$ the value $||\hat{\theta}^1_{200} - \hat{\theta}^0_{200}||$ has no meaning so we will use the convention $||\hat{\theta}^1_{200} - \hat{\theta}^0_{200}|| = 1$. The third column contains the distances of each LS estimate $\hat{\theta}^m_{200}$ from θ_0, $\left\{||\hat{\theta}^m_{200} - \theta_0|| : m = 1,2,\ldots,200\right\}$.

The estimation procedure is based on the set of distances (8) and thus, considering the values in the second column, it is easy to notice that the differences are small in the first 17 values while they increase in magnitude starting with the 18^{th} value. Such a behaviour suggests that the first 17 LS estimates $\left\{\hat{\theta}^m_{200} : m = 1,\ldots,17\right\}$ may be *close* to θ_0 and then the consistent estimate has to be chosen in that interval. Our purpose is to give theoretical support to the above intuitive arguments by introducing the following tools.

Statement 1 *The strictly convex function (3) has θ_0 as its unique global minimizer; moreover the restriction of L to each finite dimensional subspace $\mathcal{S}_m$ admits a unique global minimizer $\theta(m), \forall m \geq 1$, and $\lim_{m\to\infty} ||\theta(m) - \theta_0|| = 0$.*

Statement 2 *For any sequence of observations $\{(x_i, y_i) : i = 1,2,\ldots,n\}$ belonging to a set of probability 1, the convergences $\lim_{n\to\infty} ||\hat{\theta}^m_n - \theta(m)|| = 0$ hold true for each fixed $m \geq 1$, where $\hat{\theta}^m_n$ is the global minimizer for the random function L_n defined in (5).*

m	$\|\|\hat{\theta}^m_{200} - \hat{\theta}^{m-1}_{200}\|\|$	$\|\|\hat{\theta}^m_{200} - \theta_0\|\|$	m	$\|\|\hat{\theta}^m_{200} - \hat{\theta}^{m-1}_{200}\|\|$	$\|\|\hat{\theta}^m_{200} - \theta_0\|\|$	m	$\|\|\hat{\theta}^m_{200} - \hat{\theta}^{m-1}_{200}\|\|$	$\|\|\hat{\theta}^m_{200} - \theta_0\|\|$
1	1.00	3.31	68	1.67	236.73	135	27.07	3006.71
2	0.00	1.00	69	4.96	239.87	136	1.01	2970.53
3	0.01	0.98	70	3.27	249.06	137	65.59	2970.92
4	0.54	0.93	71	6.60	244.24	138	20.71	2984.91
5	0.00	0.56	72	8.26	258.22	139	54.03	2985.50
6	0.03	0.56	73	0.69	264.61	140	15.98	2939.51
7	0.12	0.54	74	18.81	265.55	141	4.06	2969.95
8	0.13	0.39	75	2.17	277.16	142	1844.98	2982.32
9	0.06	0.49	76	0.48	276.34	143	832.32	5119.39
10	0.36	0.61	77	0.11	279.15	144	9.18	5684.14
11	0.11	1.27	78	3.46	278.57	145	71.94	5762.81
12	0.29	1.38	79	0.85	283.09	146	508.26	5744.51
13	0.27	1.60	80	0.11	286.16	147	1.38	5417.70
14	0.21	1.75	81	2.40	285.77	148	99.36	5442.65
15	0.70	1.79	82	67.50	284.32	149	15.90	5233.33
16	0.05	2.65	83	40.29	355.73	150	15.04	5086.12
17	0.02	2.79	84	5.19	374.28	151	214.84	5170.25
18	1.21	2.86	85	5.79	382.65	152	107.44	5603.50
19	0.59	3.88	86	2.03	390.16	153	768.00	6191.55
20	1.01	4.22	87	1.37	392.82	154	511.08	7057.95
21	0.43	5.02	88	1.19	394.07	155	0.75	6787.63
22	0.08	5.01	89	28.13	395.15	156	1017.99	6755.90
23	0.71	5.15	90	126.81	428.33	157	71.41	6865.75
24	0.96	5.73	91	0.24	477.52	158	6.60	6961.57
25	1.41	6.65	92	6.73	478.64	159	87.57	6905.49
26	5.34	8.40	93	48.97	486.55	160	1298.41	6819.52
27	0.01	14.56	94	7.53	516.19	161	238.83	8596.93
28	0.02	14.53	95	0.03	518.73	162	4.24	8688.54
29	0.01	14.60	96	0.04	518.85	163	51.52	8609.61
30	4.86	14.59	97	13.54	519.22	164	0.28	9066.55
31	4.55	18.76	98	72.92	544.27	165	165.16	9082.13
32	0.09	24.50	99	130.89	567.98	166	18.38	9444.54
33	0.00	24.81	100	47.09	678.14	167	6.15	9369.37
34	0.11	24.81	101	42.20	757.71	168	666.30	9396.75
35	0.31	25.16	102	4.86	817.02	169	3095.56	9956.80
36	3.19	25.70	103	216.12	825.44	170	577.45	12240.31
37	0.59	29.00	104	37.00	1088.18	171	111.44	14103.65
38	3.23	30.30	105	41.77	1119.62	172	1293.01	14334.74
39	0.00	31.81	106	0.31	1145.05	173	230.32	17013.29
40	4.96	31.82	107	149.97	1148.42	174	52.57	16830.61
41	8.77	35.58	108	22.59	1155.19	175	53.06	16843.62
42	2.85	44.01	109	22.79	1120.24	176	1.78	16291.38
43	0.34	49.24	110	1.60	1144.63	177	1665.49	16291.96
44	4.53	50.29	111	1.56	1155.22	178	301.70	16345.65
45	2.04	54.78	112	15.98	1159.20	179	74.96	16394.48
46	0.25	58.07	113	13.17	1169.30	180	148.37	16163.95
47	1.65	59.23	114	211.14	1175.31	181	184.84	16635.95
48	0.34	60.84	115	303.41	1407.30	182	18.03	17251.91
49	0.75	60.67	116	2.84	1560.34	183	163.47	17517.19
50	0.34	60.97	117	177.83	1559.14	184	367.72	17536.16
51	22.11	60.96	118	49.97	1721.00	185	2485.26	18569.09
52	1.20	84.05	119	8.50	1798.81	186	3406.92	23138.50
53	1.32	85.47	120	114.10	1777.50	187	2029.08	24897.65
54	10.71	86.94	121	173.32	1897.45	188	2160.90	27255.93
55	3.98	90.87	122	0.42	2073.15	189	2.72	33692.87
56	0.57	101.88	123	0.95	2070.15	190	0.64	33777.13
57	1.81	101.87	124	0.69	2078.23	191	3526.11	33828.00
58	2.50	102.47	125	291.89	2070.12	192	1439.68	32032.98
59	0.00	110.30	126	1.81	2371.34	193	316.24	39053.37
60	4.44	110.32	127	4.60	2379.85	194	14358.82	41476.54
61	0.32	116.08	128	257.94	2373.78	195	26441.01	73069.02
62	38.93	115.76	129	18.92	2676.07	196	2904.56	68727.89
63	4.96	165.17	130	0.23	2773.64	197	364089.76	72849.47
64	44.23	164.34	131	285.99	2767.82	198	27892.26	567521.70
65	1.17	209.33	132	13.57	2953.44	199	64238.01	507066.10
66	0.97	209.89	133	1.62	2975.72	-	-	-
67	24.46	211.73	134	23.29	2980.22	-	-	-

Table 1: Simulations of example 1 for $\theta_0 = \ln|t-1|$

Let us apply the above statements to our problem. When n is *big enough*, the LS estimates $\hat{\theta}_n^m$ will be close to the respective limits $\theta(m)$ and this occurs uniformly for each m belonging to a finite subset. Then there exists a value $m_0(n)$ such that

$$||\hat{\theta}_n^m - \theta(m)|| \simeq 0, \qquad \text{for each } m = 1, 2, \ldots, m_0(n) \tag{9}$$

where $m_0(n)$ is a given natural satisfying $1 < m_0(n) < n$.

As a direct consequence of (9) we have

$$||\hat{\theta}_n^m - \hat{\theta}_n^{m'}|| \simeq ||\theta(m) - \theta(m')|| \tag{10}$$

for each pair $m, m' \in \mathbb{N}$ satisfying $1 \leq m, m' \leq m_0(n)$. Therefore, the idea behind the estimation of $\hat{\theta}_n$ is the existence of a natural $m_0(n)$ such that the LS finite dimensional estimates $\left\{\hat{\theta}_n^m : m = 1, \ldots, m_0(n)\right\}$ approximate closely their respective limits $\{\theta(m) : m = 1, \ldots, m_0(n)\}$ due to Statement 2.

Then, since $\theta(m) \to \theta_0$, the LS estimates $\left\{\hat{\theta}_n^m : m = 1, \ldots, m_0(n)\right\}$ also converge, thus explaining the low values taken by the first distances $||\hat{\theta}_{200}^m - \hat{\theta}_{200}^{m-1}||$ in the second column of Table 1. The value of $m_0(n)$, as well as θ_0, is completely unknown; nevertheless, when n tends to infinity it is possible to approximate it in a satisfactory way.

Since the estimation is based on the distances $||\hat{\theta}_n^m - \hat{\theta}_n^{m-1}||$ it is natural to introduce the notation used for intervals in order to denote the finite sequences of LS estimates.

Definition 1 *If a and b are natural numbers (with $a < b$) it is intuitive to denote the intervals of natural numbers $[a, b]$ as follows:*

$$[a, b] = \{m \in N : a \leq m \leq b\}. \tag{11}$$

Since our analysis is based on the sequences of consecutive finite dimensional LS estimates $\hat{\theta}_n^m, \hat{\theta}_n^{m+1}, \ldots, \hat{\theta}_n^{m+k}$ *it is useful to introduce the* intervals of LS estimates

$$[a, b] = \{\theta_n^m : a \leq m \leq b\}, \tag{12}$$

where a, b *are naturals not greater than the fixed* n*. Thus, hereafter* $[a, b]$ *will be used to denote an interval of LS estimates. For a given* $[a, b]$ *there are two quantities which characterize such a set*

$$d(n)[a, b] = \max\left\{||\hat{\theta}_n^m - \hat{\theta}_n^{m-1}|| : m = a, a+1, \ldots, b\right\} \tag{13}$$

and

$$\nu(n)[a, b] = b - a + 1, \tag{14}$$

which denote respectively the maximum distance of consecutive estimates and the cardinality of $[a, b]$*.*

Furthermore, we define as complete *a set* $[a, b]$ *satisfying the following inequalities*

$$||\hat{\theta}_n^{b+1} - \hat{\theta}_n^b|| > d(n)[a, b] \qquad ||\hat{\theta}_n^{a-1} - \hat{\theta}_n^{a-2}|| > d(n)[a, b].$$

Let us consider, for example, the set $[6, 8] = \left\{\hat{\theta}_{200}^6, \hat{\theta}_{200}^7, \hat{\theta}_{200}^8\right\}$ from the second columns of Table 1. We take the distances

$$\left\{||\hat{\theta}_{200}^6 - \hat{\theta}_{200}^5|| = 0.03; \quad ||\hat{\theta}_{200}^7 - \hat{\theta}_{200}^6|| = 0.12; \quad ||\hat{\theta}_{200}^8 - \hat{\theta}_{200}^7|| = 0.13\right\}$$

and then $d(200)[6, 8] = \max\{0.03; 0.12; 0.13\} = 0.13$, $\nu(200)[6, 8] = 3$. Since the fifth value in the second column is 0.00 and the ninth one is 0.06, the distances $||\hat{\theta}_{200}^5 - \hat{\theta}_{200}^4|| = 0.00$ and $||\hat{\theta}_{200}^9 - \hat{\theta}_{200}^8|| = 0.06$ are both smaller than $d(200)[6, 8] = 0.13$ so $[6, 8]$ is not complete.

Subsequently, we will consider only complete sets $[a, b]$, for example the completed version of set $[6, 8]$:

$$\begin{aligned}
[5, 9] =& \left\{\hat{\theta}^5_{200}, \hat{\theta}^6_{200}, \hat{\theta}^7_{200}, \hat{\theta}^8_{200}, \hat{\theta}^9_{200}\right\}, \\
d(200)[5, 9] =& \max\{0.00; 0.03; 0.12; 0.13; 0.06\} = 0.13 \\
\nu(200)[5, 9] =& 5.
\end{aligned}$$

Definition 2 *Given the two sets of LS estimates* $[a, b]$ *and* $[c, d]$*, we say that* $[c, d]$ *is* preferable *to* $[a, b]$ *if both of the following inequalities are satisfied*

$$\begin{aligned}
& d(n)[c, d] \leq d(n)[a, b] \\
& \nu(n)[c, d] = d - c + 1 \geq \nu(n)[a, b] = b - a + 1.
\end{aligned} \tag{15}$$

The concept of *preferable* set has an intuitive meaning: the most interesting finite sequence of LS estimates $\hat{\theta}^m_n, \hat{\theta}^{m+1}_n, \ldots, \hat{\theta}^{m+k}_n$ have small distances $||\hat{\theta}^m_n - \hat{\theta}^{m-1}_n||$ and possibly a high number of elements because we are looking for elements which are convergent to the unknown θ_0. For each assigned sequence of observations $\{(x_i, y_i) : i \geq 1\}$ belonging to a set with probability 1 (see Statement 2), there exists a sequence of values

$$\{m_0(n), m_1(n), d(n)[m_1(n), m_0(n)] : n \geq 1\} \tag{16}$$

where the two natural numbers $m_0(n), m_1(n)$ with $m_1(n) < m_0(n)$ define the interval of estimates

$$[m_1(n), m_0(n)] = \left\{\hat{\theta}^m_n : m_1(n) \leq m \leq m_0(n)\right\}$$

which is characterized by the value $d(n)[m_1(n), m_0(n)]$.

Furthermore, a sequence (16) exists in such a way that the following properties are satisfied

R1 $m_1(n) < m_0(n), \forall n$

R2 $\lim_{n\to\infty} m_1(n) = \lim_{n\to\infty} m_0(n) = +\infty$

R3 $\lim_{n\to\infty} [m_0(n) - m_1(n)] = \infty$

R4 $\lim_{n\to\infty} d(n)[m_1(n), m_0(n)] = 0$ and $\lim_{n\to\infty} \nu(n)[m_1(n), m_0(n)] = +\infty$

R5 $\{m_1(n)\}$ and $\{m_0(n)\}$ are both monotone not decreasing sequences.

R6 $m_1(n) < m_0(n) - m_1(n) \quad \forall n$

The sequence of random variables $\{V_n\}$ introduced below plays an important role in the approximation of values $m_0(n)$ and then in the estimation procedure.

Definition 3 *For a fixed n and any assigned interval $[m_1(n), m_0(n)]$ let V_n be the random variable*

$$V_n = \max\{b - a : [a, b] \subset [m_0(n) + 1, n] \text{ and } ||\hat{\theta}_n^m - \hat{\theta}_n^{m-1}|| \leq d(n)[m_1(n), m_0(n)] \quad \forall m = a, a + 1, \ldots, b\}. \tag{17}$$

The meaning of V_n is strictly connected with that of preferable set; in fact V_n indicates the maximum length of the intervals $[a, b] \subset [m_0(n) + 1, n]$ whose consecutive estimates have a distance not greater than $d(n)[m_1(n), m_0(n)]$.

If $V_n \geq m_0(n) - m_1(n)$, then there exists an interval $[a, b]$ *on the right-hand side* which is preferable to $[m_1(n), m_0(n)]$ and the behaviour of the $\hat{\theta}_n^m$'s is inconsistent.

Conversely, if $V_n < m_0(n) - m_1(n)$ we may state that $[m_1(n), m_0(n)]$ admits no preferable set; a preferable set $[a, b] \subset [m_0(n) + 1, n]$ cannot exist *on the left-hand side* because of the inequality R6.

The asymptotic behaviour of the sequence $\{V_n\}$ is illustrated in the following Statement.

Statement 3 *If, for each sequence of observations* $\{(x_i, y_i) : i \geq 1\}$ *belonging to a set with probability 1, a sequence of values (16) is fixed in such a way that properties R1-R6 are satisfied, then there exist two naturals* $\tilde{n}$ *and* $\tilde{k}$ *(depending on the sequence of observations) such that*

$$V_n \leq \tilde{k} \qquad \forall n \geq \tilde{n}$$

4 Definition of the estimator

In order to define the estimate $\hat{\theta}_n$ let us introduce the following finite and increasing sequence of values

$$\begin{aligned} r_1 = &\min\{r \geq 0 : \exists [a,b] \subset [1,n] \text{ satisfying } d(n)[a,b] = r, \\ &\text{and such that } [a,b] \text{ has no preferable sets}\} \\ r_2 = &\min\{r > r_1 : \exists [a,b] \subset [1,n] \text{ satisfying } d(n)[a,b] = r \\ &\text{and such that } [a,b] \text{ has no preferable sets}\} \end{aligned}$$

and so on for $r_3, r_4, \ldots$, in such a way that for $u \geq 1$

$$\begin{aligned} r_u = &\min\{r > r_{u-1} : \exists [a,b] \subset [1,n] \text{ satisfying } d(n)[a,b] = r \\ &\text{and such that } [a,b] \text{ has no preferable sets}\} \end{aligned} \tag{18}$$

Let $\arg(r_u)$ be the interval $[a,b]$ such that $d(n)[a,b] = r_u$, let

$$r_s = \min\{r_u : \arg(r_{u'}) \supset \arg(r_{u'-1}), \forall u' \geq u+1\} \tag{19}$$

$r_1 = 0.01,$	$\arg(0.01) = [m = 2, m = 3]$
$r_2 = 0.02$	$\arg(0.02) = [m = 27, m = 29]$
$r_3 = 0.13$	$\arg(0.13) = [m = 5, m = 9]$
$r_4 = 0.36$	$\arg(0.36) = [m = 5, m = 14]$
$r_5 = 0.54$	$\arg(0.54) = [m = 2, m = 14]$
$r_6 = 0.70$	$\arg(0.70) = [m = 2, m = 17]$
$\dots$	$\dots$

Table 2: Some values of r_u and $\arg(r_u)$ for example 1

and let $\arg(r_s)$ be the interval of finite dimensional LS estimates where we choose our estimate $\hat{\theta}_n$. Then we define $\hat{\theta}_n$ as follows.

Definition 4 *We define as $\hat{\theta}_n$ any element $\hat{\theta}_n^m \in \arg(r_s)$ which minimizes the set of the distances*

$$\{||\hat{\theta}_n^m - \hat{\theta}_n^{m-1}|| \quad \forall \hat{\theta}_n^m \in \arg(r_s)\}. \tag{20}$$

4.1 Examples

Consider again the simulation concerning $\theta_0 = \ln|t-1|$ reported in Table 1; we are interested in computing $\hat{\theta}_{200}$ using the elements given in (18) and (19) as well as the distances available in the second columns of each block in Table 1.

First of all, we have the distance 0.00 for $m = 2, 5, 33, 39, 59$. Following Definition 1, each of the five intervals containing only one point, [2], [5], [33], [39] and [59], has four preferable intervals. So, we cannot write $r_1 = 0.00$. In fact $r_1 = 0.01$ is the minimum value such that $\arg(r_1)$ has no preferable intervals and $\arg(0.01) = [m = 2, m = 3]$; see some other values in Table 2.

It is easy to check that $\arg(r_3) \subset \arg(r_4) \subset \arg(r_5) \subset \arg(r_6) \subset \dots$ and therefore $r_s = r_3; \arg(r_3) = [m = 5, m = 9]$ and $d(200)[m = 5, m = 9] = \max\{0.00; 0.03; 0.12; 0.13\} =$

0.13. Furthermore, applying the definition, we obtain

$$\hat{\theta}_{200} = \hat{\theta}^5_{200}.$$

The distances $||\hat{\theta}^m_n - \theta_0||$ reported in the third column allow us to check the precision of our estimate: $||\hat{\theta}^5_{200} - \theta_0|| = 0.56$ and $||\hat{\theta}^8_{200} - \theta_0|| = 0.39$ where $\hat{\theta}^8_{200}$ is the optimal LS estimate, i.e. the finite dimensional estimate which minimizes the distance from θ_0. It is important to note that $\hat{\theta}^8_{200}$ is an element of r_s.

Another interesting example is given by $\theta_0(t) = \sin(4\pi t)$; a simulation for $n = 200$ is included in Table 3. As in the previous case, the estimation $\hat{\theta}_{200}$ is computed on the basis of the distances $||\hat{\theta}^m_{200} - \hat{\theta}^{m-1}_{200}||$ available in the second column. The first elements of the sequence $\{r_u\}$ are given in Table 4.

In this case, $r_s = r_2 = 0.01$ and $\arg(r_s) = [m = 20, m = 30]$ so $\hat{\theta}_{200}$ is obtained choosing any element within the set $\{\hat{\theta}^{20}_{200}, \hat{\theta}^{21}_{200}, \hat{\theta}^{23}_{200}, \hat{\theta}^{24}_{200}, \hat{\theta}^{25}_{200}, \hat{\theta}^{26}_{200}, \hat{\theta}^{28}_{200}, \hat{\theta}^{29}_{200}, \hat{\theta}^{30}_{200}\}$.

A relevant aspect emerges considering both of the above examples. On one hand, the intervals $\arg(r_s)$ contain the optimal estimate $\hat{\theta}^m_n$ which minimizes the distance with θ_0, whereas, on the other hand, considering the values in the third column, it is easy to check that immediately after $\arg(r_s)$ the distances of the estimates $\hat{\theta}^m_n$ from θ_0 assume consistently higher values.

The simulations of $\theta_0(t) = \sin(4\pi t)$ are of particular interest, as we are able to compare the estimates produced by the method presented in this paper and the results given by Cardot et al. (1999, 2003), who derived an estimator for θ_0 through a method introduced by Bosq (1991, 2000) in the case of ARH processes.

The next section is devoted to the main proofs provided in a more general fashion which is suitable for the estimation of a not necessarily linear regression function. To this

m	$\|\|\hat{\theta}^m_{200} - \hat{\theta}^{m-1}_{200}\|\|$	$\|\|\hat{\theta}^m_{200} - \theta_0\|\|$	m	$\|\|\hat{\theta}^m_{200} - \hat{\theta}^{m-1}_{200}\|\|$	$\|\|\hat{\theta}^m_{200} - \theta_0\|\|$	m	$\|\|\hat{\theta}^m_{200} - \hat{\theta}^{m-1}_{200}\|\|$	$\|\|\hat{\theta}^m_{200} - \theta_0\|\|$
1	1.00	1.58	68	0.00	0.22	135	0.78	4.75
2	0.00	1.58	69	0.05	0.22	136	0.00	6.04
3	0.00	1.58	70	0.00	0.29	137	0.27	6.06
4	0.00	1.58	71	0.01	0.30	138	0.04	6.26
5	0.00	1.58	72	0.01	0.30	139	0.44	6.11
6	0.00	1.58	73	0.00	0.32	140	0.00	6.41
7	0.00	1.59	74	0.01	0.32	141	0.47	6.51
8	0.01	1.58	75	0.01	0.33	142	0.02	7.44
9	0.02	1.60	76	0.02	0.34	143	0.07	7.53
10	0.00	1.59	77	0.05	0.35	144	0.11	7.74
11	0.27	1.58	78	0.04	0.39	145	0.88	8.24
12	1.11	1.26	79	0.05	0.41	146	0.03	9.81
13	0.12	0.23	80	0.00	0.45	147	0.82	10.22
14	0.05	0.12	81	0.25	0.45	148	0.00	10.47
15	0.04	0.09	82	0.04	0.71	149	0.05	10.49
16	0.02	0.07	83	0.00	0.74	150	0.00	10.31
17	0.01	0.06	84	0.04	0.74	151	1.51	10.35
18	0.02	0.05	85	0.05	0.75	152	1.86	13.32
19	0.02	0.05	86	0.06	0.80	153	0.02	12.70
20	0.00	0.05	87	0.07	0.79	154	0.73	13.03
21	0.00	0.05	88	0.03	0.86	155	0.07	13.72
22	0.01	0.05	89	0.00	0.89	156	0.54	13.64
23	0.00	0.04	90	0.11	0.90	157	0.82	14.94
24	0.00	0.04	91	0.06	0.89	158	1.50	16.78
25	0.00	0.04	92	0.05	0.99	159	0.00	19.71
26	0.00	0.04	93	0.03	1.00	160	0.02	19.82
27	0.01	0.04	94	0.00	1.00	161	1.88	20.00
28	0.00	0.04	95	0.00	1.01	162	0.03	17.42
29	0.00	0.04	96	0.04	1.01	163	0.51	17.83
30	0.00	0.04	97	0.02	1.09	164	0.38	20.06
31	0.02	0.04	98	0.02	1.14	165	0.02	20.55
32	0.00	0.05	99	0.02	1.18	166	0.12	20.34
33	0.00	0.05	100	0.00	1.24	167	1.52	19.57
34	0.00	0.05	101	0.00	1.24	168	0.27	17.56
35	0.02	0.05	102	0.06	1.23	169	0.39	16.52
36	0.01	0.06	103	0.16	1.35	170	1.04	17.68
37	0.00	0.06	104	0.12	1.45	171	1.38	17.27
38	0.00	0.07	105	0.00	1.42	172	0.02	20.55
39	0.00	0.07	106	0.07	1.43	173	0.37	20.62
40	0.00	0.08	107	0.00	1.48	174	0.03	21.59
41	0.00	0.08	108	0.21	1.47	175	0.94	21.42
42	0.01	0.08	109	0.00	1.79	176	1.13	24.28
43	0.00	0.08	110	0.10	1.79	177	1.81	24.82
44	0.00	0.08	111	0.02	1.87	178	0.32	26.08
45	0.02	0.08	112	0.02	1.93	179	1.46	25.05
46	0.00	0.10	113	0.39	1.97	180	0.10	22.22
47	0.01	0.10	114	0.09	2.33	181	1.02	21.90
48	0.02	0.11	115	0.57	2.35	182	9.30	24.08
49	0.00	0.13	116	0.00	3.27	183	1.73	35.10
50	0.02	0.13	117	0.00	3.26	184	5.47	37.97
51	0.01	0.15	118	0.00	3.27	185	1.04	34.35
52	0.00	0.15	119	0.28	3.28	186	5.17	36.18
53	0.00	0.15	120	0.00	3.57	187	0.06	46.45
54	0.01	0.15	121	0.09	3.58	188	0.15	47.67
55	0.01	0.15	122	0.32	3.55	189	6.76	47.78
56	0.00	0.16	123	0.01	3.85	190	9.91	40.98
57	0.00	0.16	124	0.03	3.85	191	2.92	53.17
58	0.03	0.16	125	0.12	3.88	192	6.95	46.87
59	0.01	0.18	126	0.06	4.15	193	37.81	61.84
60	0.00	0.19	127	0.04	4.28	194	0.35	133.58
61	0.00	0.20	128	0.01	4.45	195	81.28	141.93
62	0.00	0.20	129	0.09	4.40	196	67.93	141.23
63	0.01	0.20	130	0.12	4.51	197	176.96	98.04
64	0.00	0.20	131	0.03	4.74	198	791.24	320.70
65	0.02	0.20	132	0.17	4.78	199	264.20	1148.34
66	0.00	0.23	133	0.03	4.62	-	-	-
67	0.00	0.22	134	0.06	4.60	-	-	-

Table 3: Simulations of example 2 for $\theta_0 = \sin(4\pi t)$

$r_1 = 0.00,$	$\arg(0.00) = [m = 2, m = 7]$
$r_2 = 0.01,$	$\arg(0.01) = [m = 20, m = 30]$
$r_3 = 0.02,$	$\arg(0.02) = [m = 16, m = 57]$
$r_4 = 0.03,$	$\arg(0.03) = [m = 16, m = 68]$
$\ldots$	$\ldots$

Table 4: Some values of r_u and $\arg(r_u)$ for example 2

aim, we suppose that X is a random element taking values into an arbitrary complete and separable metric space $\mathcal{X}$ having the Borel σ-field $\mathcal{B}_{\mathcal{X}}$, whereas, Y is a real random variable such that

$$E(Y|X = x) = T_0(x) \tag{21}$$

where $T_0 \in L^2(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P_X)$ is an assigned square integrable function belonging to the separable Hilbert space L^2 and P_X denotes the probability measure induced by X. The analysis given below will be performed following closely the approach developed in Section 3 for the functional linear case and keeping the same notations although with the following exceptions:

i. T_0 (and not θ_0) is the unknown parameter to estimate;

ii. the set $\{\phi_j : j \geq 1\}$ is an orthonormal basis of $L^2(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P_X)$ and $\mathcal{S}_m = Sp\left(\phi_j : j = i, \ldots, m\right)$ is the finite dimensional subspace generated by the first m elements of the basis. $L(T) = E[(Y - T(X))^2]\ \forall T \in L^2(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P_X)$ is a strictly convex real function having T_0 as its unique global minimizer;

iii. $\theta(m)$ is still the global minimizer for the restriction of $L(\cdot)$ to the subspace $\mathcal{S}_m$; analogously, $\hat{\theta}_n^m$ denotes the global minimizer for the function

$$L_n(a_j : j = 1, \ldots, m) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{m} a_j \phi_j(x_i) \right)^2 ;$$

iv. the used norm is the usual L^2 norm to respect to the measure P_X, i.e.

$$||T|| = \left(\int_{\mathcal{X}} T^2(x) dP_X(x) \right)^{1/2}.$$

5 Technical results

The following two assumptions are the only hypotheses required by the following results.

A1 We assume that the conditional random variable $Y|X = x$ admits a density function $f_{Y|X=x}(y)$ satisfying the boundedness condition

$$f_{Y|X=x}(y) \leq M \qquad \forall y \in R^1, \quad \forall x \in \mathcal{X}$$

for some constant M.

A2 We assume that $Var(Y|X = x)$ is a P_X- integrable function, i.e.

$$\int_{\mathcal{X}} Var(Y|X = x) dP_X(x) < \infty$$

Using **A1** and **A2** we can prove the following formal properties of the function $L(\cdot)$:

I) $L(T)$ is finite $\forall T \in L^2(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P_X)$

II) $L(\cdot)$ is a strictly convex function having T_0 as its unique global minimizer; furthermore $L(\cdot)$ is a real continuous function over all the domain $L^2(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P_X)$ with respect to the L^2 norm $||\ ||$.

Our purpose is to prove Statement 1 using the *separation property* that we now introduce as a preliminary tool.

Definition 5 (Separation property) *The real convex function f, defined over the vector space V, and with a unique global minimizer $v_0 \in V$, is said to satisfy the* Separation Property *(S.P. hereafter) if for each fixed $\epsilon > 0$ there exists a corresponding $\delta(\epsilon) > 0$ such that for any $v \in V$ with*

$$d(v, v_0) \geq \epsilon \Rightarrow f(v) - f(v_0) \geq \delta(\epsilon)$$

where d denotes the metric defined over V.

The S.P. has an intuitive meaning; choosing any point v such that $d(v, v_0) \geq \epsilon$ the difference $f(v) - f(v_0)$ cannot be arbitrarily close to zero: in fact there exists $\delta(\epsilon) > 0$ such that $f(v) - f(v_0) \geq \delta(\epsilon)$. The S.P. was introduced by R.T. Rockafellar for a convex function defined over R^n (see Theorem 27.2 on page 265 in Rockafellar, 1972). In the case of a function defined over an infinite dimensional vector space, the S.P. is not easy to obtain. Nevertheless, in the particular case of the strictly convex function $L(\cdot)$ having T_0 as its unique global minimizer, the S.P. holds true over all the domain.

Lemma 1 *Given the function, $L(\cdot)$ for each fixed $\epsilon > 0$ there exists a corresponding value $\delta(\epsilon) > 0$ such that for all $T \in L^2(\mathcal{X}, \mathcal{B}_\mathcal{X}, P_X)$ satisfying $||T - T_0|| \geq \epsilon \Rightarrow L(T) - L(T_0) \geq \delta(\epsilon)$.*

Sketch of the proof. Given the half-line having its origin in T_0

$$T_0 + t(T - T_0) \qquad \forall t \geq 0$$

and adopting as the first derivative

$$L'(t) = 2t \int_\mathcal{X} (T(x) - T_0(x))^2 dP_X(x) = 2t||T - T_0||^2$$

we have that $L(\cdot)$ has the same behaviour over each half-line.

Theorem 1 (Statement 1) *The strictly convex function* $L(T) = E[(Y - T(X))^2]$, $\forall T \in L^2(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P_X)$, *has* T_0 *as its unique global minimizer; moreover the restriction of* L *to each finite dimensional subspace* $\mathcal{S}_m$ *admits a unique global minimizer* $\theta(m)$, $\forall m \geq 1$ *and* $\lim_{m\to\infty} ||\theta(m) - T_0|| = 0$.

Proof. The proof that T_0 is the unique minimizer for L is omitted and then we consider the existence and convergence for the sequence $\{\theta(m)\}$.

For each natural m fixed let us denote by $T_m = \sum_{j=1}^{m} a_j^0 \phi_j$ the m-th term of the Fourier series of T_0, where a_j^0 are the Fourier coefficients. Then $\lim_{m\to\infty} ||T_m - T_0|| = 0$ and there exists a real $\tau > 0$ such that

$$\bigcup_{m=1}^{\infty} T_m \subset \bar{S}(T_0, \tau)$$

in such a way that $\mathcal{S}_m \cap \bar{S}(T_0, \tau) \neq \emptyset$. Let us consider its closure set denoted by $cl[\mathcal{S}_m \cap \bar{S}(T_0, \tau)]$; such a set is a closed and bounded subset of a finite dimensional vector space and, because of the continuity and the strict convexity of L, there is a unique minimizer $\theta(m)$ of L, over $cl[\mathcal{S}_m \cap \bar{S}(T_0, \tau)]$. Note that for a given $T \in cl[\mathcal{S}_m \cap \bar{S}(T_0, \tau)]$ satisfying $||T - T_m|| \geq \epsilon$ we have, due to the triangular inequality, that

$$||T - T_0|| \geq ||T - T_m|| - ||T_m - T_0|| \geq \epsilon - ||T_m - T_0|| \qquad \forall m \text{ fixed;}$$

when $m \to \infty$ we have that $||T_m - T_0|| \to 0$ while ϵ is a fixed constant and then

$$||T - T_m|| \geq \epsilon \Rightarrow ||T - T_0|| \geq \epsilon - ||T_m - T_0|| > 0$$

for m *big enough.* The strict positivity of the difference $\epsilon - ||T_m - T_0||$ allows us to apply

the S.P. and then the existence is stated for a value $\delta(\epsilon - ||T_m - T_0||)$ such that

$$L(T) - L(T_0) \geq \delta(\epsilon - ||T_m - T_0||) \tag{22}$$

We consider now the difference $L(T) - L(T_m)$ when $T \in cl[\mathcal{S}_m \cap \bar{S}(T_0, \tau)]$ and $||T - T_m|| \geq \epsilon$:

$$L(T) - L(T_m) = L(T) - L(T_0) + L(T_0) - L(T_m) = (L(T) - L(T_0)) - (L(T_m) - L(T_0)).$$

Applying the inequality (22) to the difference $L(T) - L(T_0)$ we have that

$$L(T) - L(T_m) = (L(T) - L(T_0)) - (L(T_m) - L(T_0)) \geq \delta(\epsilon - ||T_m - T_0||) - (L(T_m) - L(T_0)).$$

If $m \to \infty$ it follows that $||T_m - T_0|| \to 0$ and $(L(T_m) - L(T_0)) \to 0$ because of the continuity of L, while $\epsilon - ||T_m - T_0|| \to \epsilon$ and then we have, when m is *big enough*, that

$$||T - T_m|| \geq \epsilon \Rightarrow L(T) - L(T_m) \geq \delta(\epsilon - ||T_m - T_0||) - (L(T_m) - L(T_0)) > 0.$$

Taking now the minimizer $\theta(m)$ for L over $cl[\mathcal{S}_m \cap \bar{S}(T_0, \tau)]$, we have necessarily that $L(\theta(m)) - L(T_m) \leq 0$ and the equality holds true when $\theta(m) = T_m$. Then $||\theta(m) - T_m|| < \epsilon$ and

$$||\theta(m) - T_0|| \leq ||\theta(m) - T_m|| + ||T_m - T_0|| \leq \epsilon + ||T_m - T_0||.$$

If we choose an arbitrarily small $\eta > 0$ there exists $m(\eta)$ such that

$$||\theta(m) - T_0|| \leq \epsilon + \eta.$$

Finally the strict convexity of L allows us to prove that each $\theta(m)$ is the global minimizer for L over the space $\mathcal{S}_m$.

Theorem 2 (Statement 2) *For any sequence of observations $\{(x_i, y_i) : i \geq 1\}$, belonging to a set of probability one, the following statements hold true*

- *for each fixed m the sequence of convex random functions*

$$L_n(a_j : j = 1, \ldots, m) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{m} a_j \phi_j(x_i) \right)^2$$

is convergent to the strictly convex function

$$L(a_j : j = 1, \ldots, m) = E \left[\left(Y_i - \sum_{j=1}^{m} a_j \phi_j(X) \right)^2 \right]$$

uniformly over each compact subset $K \subset R^m$;

- *for each fixed m, $\lim_{n\to\infty} ||\hat{\theta}_n^m - \theta(m)|| = 0$ where $\hat{\theta}_n^m$ and $\theta(m)$ denote respectively the global minimizer for L_n and L over the m-dimensional subspace $\mathcal{S}_m$.*

Proof. Applying countably many times the strong law of large numbers the convergence is obtained

$$\lim_{n\to\infty} L_n(a_j : j = 1, \ldots, m) = L(a_j : j = 1, \ldots, m)$$

for each point $(a_j : j = 1, \ldots, m)$ belonging to a dense and countable subset of R^m. The first comma is then proved by applying the Theorem 10.8 on page 90 of Rockafellar (1972). Furthermore, as the limit function L is strictly convex and with a unique global minimizer $\theta(m)$ over $\mathcal{S}_m$, the S.P. holds true for L. Thus for fixed $\epsilon > 0$, there exists a positive $\delta(\epsilon)$ such that for each point $(a_j : j = 1, \ldots, m)$ not belonging to the compact ball $\bar{S}(\theta(m), \epsilon)$ we have

$$L(a_j : j = 1, \ldots, m) - L(\theta(m)) \geq \delta(\epsilon).$$

Finally, because of the convergence of L_n to L uniformly over the compact set $\bar{S}(\theta(m), \epsilon)$ it is easy to prove that $\hat{\theta}_n^m$ belongs to $\bar{S}(\theta(m), \epsilon)$ for n big enough.

6 Conclusions and remarks

Lastly we observe that the strong consistency of $\hat{\theta}_n$ may be obtained via Statement 3 proving that for each sequence of observations $\{(x_i, y_i) : i \geq 1\}$ belonging to a set with probability one there exists $\bar{n}$ (depending on the given sequence of observations) such that the inclusion

$$\arg(r_s) \subset [m_1(n), m_0(n)] \qquad \forall n > \bar{n}$$

is satisfied.

The proposed method shows interesting results also in estimating not regular functions; for instance, several other simulations show the possibility of detecting the position of discontinuities as well as the jumps size of a regression function.

References

Besse, P., Cardot, H., and Stephenson, D. (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journ. of Statist.* 27:673–687.

Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In *Roussas, G. (Ed.), Nonparametric Functional Estimation and Related Topics*, pages 509–529. NATO, ASI series.

Bosq, D. (2000). *Linear processes in function spaces.* Number 149 in Lecture Notes in Statistics. Springer Verlag, New York.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45:11–22.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, 13:571–591.

Cardot, H. and Johannes, J. (2010). Thresholding projection estimates in the functional linear model. *Journal of Multivariate Analysis*, 101:395–408.

Cardot, H., Crambes, C., Kneip, A. and Sarda, P. (2007). Smoothing splines estimators in functional linear regression with errors-in-variables. *Comput. Statist. & Data Anal.* 51:4832–4848.

Frank, I. and Friedman, J. (1993). A Statistical view of some chemiometrics regression tools. *Technometrics* 35:10–148.

Müller, H.G. and Stadtmüller, U. (2005). Generalized functional linear models *Ann. Statist.* 33:774-805.

Preda, C. and Saporta, G. (2005). PLS regression on a stochastic process. *Comput. Statist. & Data Anal.* 48:149–158.

Reiss, P.T., and Ogden, R.T. (2007). Functional principal component regression and functional partial least squares. *J.A.S.A.*, 102(479):984–996.

Rockafellar, R. (1972). *Convex Analysis*. Princeton University Press.